DOCUMENT RESUME

ED 481 653 TM 035 354

AUTHOR	Yu, Lei; White, Donald B.
TITLE	Measuring Value Added School Effects on Ohio Six-Grade Proficiency Test Results Using Two-Level Hierarchical Linear Modeling.
PUB DATE	2002-04-00
NOTE	45p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 2002).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Achievement Tests; *Elementary School Students; Intermediate Grades; Measurement Techniques; *School Effectiveness; State Programs; *Testing Programs
IDENTIFIERS	*Ohio Sixth Grade Proficiency Test; Testing Effects; *Value Added

ABSTRACT

This study was conducted to measure value added school effects in a Northwest urban public school district using a two-level hierarchical model. The model consisted of two student-level variables (prior achievement and eligibility for federal free or reduced-price lunch) and three school variables (percentage of students eligible for free or reduced-price lunch, percentage of teachers with a Master's degree, and percentage of students suspended). The outcome measure was composite scores derived from the results of the Ohio Proficiency Test on five subject areas: writing, reading, citizenship, mathematics, and science). The sample included 1,915 sixth graders (1999-2000 school year), in 44 elementary schools. Findings show that schools can and do make a difference to student academic achievement and that school effects are not uniform across students with different prior achievement. (Contains 8 tables and 53 references.) (Author/SLD)

Measuring Value Added School Effects on Ohio Sixth-Grade Proficiency Test Results Using Two-Level Hierarchical Linear Modeling

Lei Yu

Donald B. White

University of Toledo

Paper Presented at the Annual Meeting

at the American Educational Research Association

New Orleans, LA

April, 2002

Abstract

This study is conducted to measure value added school effects in a Northwestern urban public school district using a two-level hierarchical linear model. The model consists of two student level variables (prior achievement and eligibility to federal free- or reduced-price lunch) and three school level variables (percentage of students eligible to free- or reduced-price lunch, percentage of teachers with a Master's degree, and percentage of students suspended). The outcome measure is composite scores derived from the results of Ohio Proficiency Test on five subject areas: writing, reading, citizenship, math, and science. The sample includes 1915 sixth-grade students (1999-2000 school year), who are hosted in the 44 elementary schools. The study reveals that schools can and do make a difference to student academic achievement and that school effects are not uniform across students with different prior achievement.

Introduction

Researchers have been studying schools' efforts to bring about educational outcomes on their students since the mid-sixties (Cohen, 1983; Goldstein, 1995; Holdaway & Johnson, 1993; Mandeville & Heidari, 1988). Research in this area is often conducted to measure the existing level of effectiveness of educational organizations and programs (Holdaway & Johnson, 1993), to compare schools in terms of their effects on student achievement (Goldstein, 1995), to identify effective or exceptional schools (Frederick, 1987; Scheerens & Bosker, 1997; Webster, Mendro, & Almaguer, 1993), or to investigate effective school characteristics (Brown, Fuffield, & Riddell, 1997; Goldstein, 1995, 1997; Creemers, 1994). With the increasing public concern over the effectiveness of the use of educational resources and growing demand to hold schools accountable for their influence on educational outcomes, school accountability becomes another focus of school effectiveness research (Cohen, 1983; Caldas & Bankston III, 1997; Frederick, 1987; Mandeville & Heidari, 1988).

The most important finding of school effective research is that schools make a difference (Good & Weinstein, 1986; Odden & Webb, 1983; Thrupp, 1999; Witte & Walsh, 1990). Studies show that the schools students attend have a substantial effect on their educational progress (Good & Weinstein, 1986) and the amount of learning varies from school to school (Good & Weinstein, 1986; Odden & Webb, 1983), even when taking into account their prior attainment and background characteristics (Odden & Webb, 1983). Studies also show that the amount of variability in students' academic achievement that can be explained by school- and teacher-level factors range from 10 - 18%. Brown et al. (1997) noted that education accounted for, at most, 10 - 15% of the variance in student achievement and that this amount was considerably less than what could be explained by family and community background variables. Creemers (1994) summarized the results of school effectiveness research and concluded that school and classroom factors explained 12 to 18% of the variance in student achievement after taking student background into account. In their sample of U.S. studies using multilevel modeling, Bosker and Witziers (1996) reported that school level factors accounted for 10% of the variance in student achievement after adjusted for student background.

One of the characteristics of schooling is that students are nested within classes and classes are nested within schools (McPherson, 1997). This nested structure dictates the hierarchical nature of the data. Hierarchical linear modeling (HLM), also known as multilevel modeling, is now an "established technique" with growing applications in SER (Goldstein, 1997). As an extension of regression, HLM incorporates both student- and school-level variables to explain within- and between-school variation in student achievement scores over and above the predictors at each level (Phillips & Adcock, 1997; Saunders, 1997). Among its many applications, the model has been utilized to rank schools on their effect to improve student achievement or to study factors related to school effectiveness through the use of an effective indicator as the dependent variable (Goldstein, 1997; Phillips & Adcock, 1997). Its great success lies in unraveling the nature of schooling process more accurately and more adequately than traditional regression analysis (Schagen, 1991; Reynolds & Teddlie, 2000). Hierarchical linear modeling has now become "the standard method for reporting the results of school effects" (Teddlie, Reynolds, & Sammons, 2000, p. 113).

The most recent development in the field is the value added definition of school effects. As Stoll & Fink (1996) quoted from Mortimore, "an effective school is one in which pupils progress further than might be expected from consideration of its intake" (p. 27). This, put simply, refers to value added, which describes "the boost" that a school gives to a student's prior achievement, after controlling for prior achievement and background factors (McPherson, 1997; Stoll & Fink, 1996). This conceptualization is aimed at leveling the playing field for those students who are in a disadvantaged academic position because of their family or community environments (Stoll & Fink, 1996). Value added measure reveals that schools can and do make a difference on student achievement (Sammons, Mortimore, & Thomas, 1996; Saunders, 1997). This notion of school effects is followed in this study, which is operationally defined as the differences among schools in terms of their ability to influence student academic achievement, controlling for student intake characteristics (e.g., prior achievement) and school-level factors (e.g., school SES and disciplinary climate).

The value added measure means that the criterion of effectiveness is an outcome measure over and above student intake characteristics, including prior attainment and background factors (Sheerens & Bosker, 1997). Value added is different from raw scores and adjusted scores. Raw scores are the achievement or test scores reported by testing agencies while adjusted scores are scores adjusted for student background characteristics such as age, gender, and ethnicity (Sammons et al., 1996). Value added takes into account student prior achievement besides background characteristics. Raw score represents the actual level of achievement, but they do not measure the value added by a school (McPherson, 1997).

A lot of research has been conducted to find out factors contributing to student academic achievement. The relationship between students' family background characteristics and academic achievement is well established (Willms, 1992). It has been found that social-economic status (SES) has an enduring effect on achievement even when controlling for all kinds of input variables (Caldas, 1993). Saunders (1997) found that low SES has a strong negative correlation with all educational outcomes. Research also shows that the SES composition of student body affects achievement, independent of individual student's own family background. Attending school with classmates from families of higher SES increases one's own academic performance regardless of one's own SES status, race and other factors (Caldas & Bankston III, 1997; Goldstein, 1997). Orderly and safe environment has been consistently reported to be a characteristic of unusually effective schools (Purkey & Smith, 1983, Austin & Reynolds, 1990), but it may not be a factor that differentiates between more and less effective schools (Levine & Lezotte, 1990). ETS (1999) conducted a study using the National Educational Longitudinal Study of 1988 (NELS: 88) to measure the relationship between discipline and achievement. It was found that discipline problems were associated with low achievement. Lower levels of discipline problems were associated with higher levels of achievement, and higher levels of discipline problems were associated with lower levels of achievement.

The effects of teachers on student achievement have received considerable attention in recent years. Substantial evidence from studies conducted in Tennessee, Dallas, and Boston demonstrates that teachers make a difference in student achievement (Haycock, 1998; The Thomas B. Fordham Foundation, 1999) and that teacher quality is crucial to academic success (Stone, 1999). Sanders and Rivers (1996) examined cumulative teacher effects in mathematics (grades 3 to 5) in two large city school systems in Tennessee. The teachers were classified into five quintiles according to their degree of effectiveness estimated by a statistical mixed model and then linked to the students they taught. The results show that students who had similar achievement levels in grade 2 and were taught by the least effective and most effective teachers for three years resulted in an averaged difference of 52 to 54 percentile points in math achievement. A few other studies also show that teacher effects, instead of classroom contextual variables or class size, are a dominant factor that influences student achievement (Sanders, 1998; Sanders, Wright and Horn, 1997).

On the basis of Sanders and Rivers' research (1996), Mendro, Jordan, Gomez, Anderson and Bembry (1998) conducted a similar study examining longitudinal teacher effectiveness using a hierarchical linear model. They found that, on average, grade 4 students taught by less effective teachers for three years had a percentile rank in reading scores lowered from 60 to 21, while those taught by more effective teachers had a percentile rank raised from 56 to 59. Hanushek (1992) noted that "the estimated difference in annual achievement growth between having a good and having a bad teacher can be more than one grade-level equivalent in test performance" (p. 107).

Purpose of the Study

For school improvement and academic accountability, the Ohio Senate Bill 55, passed by the Ohio General Assembly in August 1997, established a performance accountability rating system to evaluate school and districts based on their performance (Ohio Department of Education, 2000). School districts are rated against the number of

State performance standards and assigned to one of the four ratings: effective, continuous improvement, academic watch, or academic emergency (Ohio Department of Education, 2000). This rating system has a lot of advantages, such as monitoring the real level of student academic achievement and assessing a district's performance relative to an external benchmark. However, this may be not a fair rating system. Although a district's standing as compared to "similar districts" is provided, similar districts are established on five factors, including district size, poverty level, SES, factors related to urban or rural location and overall property wealth. No prior achievement measure is included. Schools are also assessed and issued local report cards using similar State performance standards. A school's performance as compared to the District average and State average are also provided.

The accountability system that compares schools and districts using unadjusted outcome measures "favors schools that serve advantaged students and usually adversely affect schools with population demographics that differ from the norm" (Webster et al., 1993, p. 4). Schools of higher SES will typically perform better than schools of lower SES on these unadjusted measures (Sammons et al., 1996). In a recent study conducted on comparing the most advantaged and the least advantaged school districts in Ohio on the State 2000 Report Card designation, Rachor (2000) found that 28% of the schools in advantaged districts, as compared to 0% of those in the disadvantaged districts, were rated as effective and that 0% of the schools in advantaged districts, as compared to 45% of those in the disadvantaged districts, were rated academic emergency. An education accountability system that fails to consider gains relative to prior achievement leads to "misleadingly negative evaluations for educators who are producing substantial but insufficient gains with disadvantaged students or misleadingly positive evaluations of educators who are producing mediocre gains with talented and advantaged students" (Stone, 1999, p. 243).

It is generally agreed that scores, if not adjusted for intake characteristics, do not measure school effectiveness (Coe & Fitz-Gibbon, 1998; Goldstein, 1997; Strand, 1997). Value-added measure can be used to "appraise fairly and accurately school and system performance regardless of differences among entering students" (Stone, 1999, p. 240). Research on value added measure of school effects using Ohio Proficiency Test results is very limited. This study is designed to measure how much elementary schools vary in their students' educational outcomes after taking into account student intake characteristics (e.g., student prior achievement and background characteristics) and school level factors (e.g., school SES and disciplinary climate). The research question is: "To what extent do schools vary in the academic achievement of their students?" To be specific, the purpose of the study is to estimate value-added school effects and rank order schools on their estimates of school effects. School effects refer to the ability of schools to affect or modify student achievement (Raudenbush & Willms, 1995; Teddlie, Reynolds, & Sammons, 2000). In this study, they are operationally defined as the differences among schools in terms of their ability to influence student academic achievement, after controlling for student intake characteristics and school-level factors.

Method

Sample

This study was designed to measure school effects on Ohio Sixth-Grade Proficiency Test (OPT) scores. Ohio Proficiency Test covers five subject areas: reading, writing, mathematics, science, and citizenship. Principal Components Analysis (PCA) was used to derive a composite score as the outcome measure representing students' overall academic achievement. The subjects included in the study were grade 6 students (1999-2000 school year) that were served in all the elementary schools in a mid-western urban public school district. The total number of schools was 44. These schools were heterogeneous with regard to community type, size, and socioeconomic status (SES). Sixth-grade students in regular education classes were considered as eligible students and those with disabilities were excluded.

The total number of eligible sixth-grade students in the district was 2328. Among these students, 409 were excluded from the study, as they did not have any valid fourth-grade Ohio Proficiency Test scores. This resulted in a sample of 1919 students, 1785 of which had complete data on all the tests and 134 of which had missing observations on at least one portion of the sixth-grade or fourth-grade Ohio Proficiency Tests. Missing values for the 134 cases were imputed. The imputation procedure was explained in the next section. Four cases were dropped for the reason that the variance of imputed factor scores was relatively large. The final sample size was 1915 students from 44 schools. All the data used in the study were obtained from district's research department, which maintains and manages all the databases.

Treatment of Missing Values

To impute the missing values, the technical procedure illustrated by Johnson and Wichern (1998) was followed, modified, and applied to the data requirement for this study. Imputation consisted of the following steps. First, the population mean and variance/covariance were estimated using statistics from complete data (cases with observed scores on all of the five content areas of the OPT). Second, the obtained estimates were then used to predict the conditional means given available data for a single case. Third, individual missing values equal to these conditional means were imputed and then used to impute the factor score. Fourth, the conditional variance and covariance of the imputed values were estimated and used to calculate the variance of the imputed factor scores. The estimated variance of the imputed factor score was used to judge the quality of imputed values and determine whether they should be retained or dropped. The advantage of this method is that it provides an intermediate check on the accuracy of imputed values via variance estimate. If variance is small, the imputed values are trustworthy. If variance is too large, their inclusion in the analysis is problematic. The result shows that the variance of the imputed factor score was large when there were more than two missing observations for a case. Based upon the distributions of standard deviations for the imputed factor scores, imputed scores with a standard deviation less or equal to .5 were kept and the cases with that greater than .5 were dropped. With the adding in of the imputed values, scores were then standardized with a mean of 100 and standard deviation of 15.

Variables

As mentioned above, the outcome measure used in this study was the composite score derived from five areas covered by Ohio Proficiency Test. Research shows that composite scores are more stable than subcomposite or component scores in measuring school effects (see Crone et al., 1994). Principal components analysis (PCA) was used to create a composite score out of these five scores as an indicator of the overall outcome measure for each student. Principal components analysis is often used to reduce a complex set of correlated variables to a smaller and more manageable set and make interpretation simpler. Such a process results in one or more factors or components that will account for most of the variance in the original variables (Stevens, 1996). The subjects' corresponding fourth grade OPT scores were used as prior achievement measure (PRIORACH), which was manipulated in the same way as the outcome measure. These composite scores were then standardized with a mean of 100 and standard deviation of 15. Another student level variable is socioeconomic status, measured by eligibility to federal free or reduced-price lunch (LUNCH). This measure is often used as an indicator of family income since only students from low-income families are eligible to free- or reduced-price lunch.

Three school level variables were included in the study. They were percentage of students in a school eligible for federal free- or reduced-price lunch (LUNCH%), disciplinary climate, which is measured by percentage of students suspended (SUSP%), and percentage of teachers with a Master's degree (TM%).

Analysis

The study was conducted using two-level hierarchical linear modeling. The model consists of two levels: level 1 being the student level and level 2 being school level. Level 1 model specifies relationships among student-level variables and the outcome is represented as a function of individual student characteristics. This is the standard linear model except that the regression coefficients are allowed to vary across schools. The regression coefficients are then conceived as an outcome variable at the school level. Their variability is estimated and modeled as a function of school characteristics. The estimation of school effects was conducted following a three-step procedure:

unconditional model, random coefficient model, and intercept- and slopes-as-outcomes model (Bryk & Raudenbush, 1992; Lee, 2000).

Unconditional model. The unconditional model did not include any predictors at either Level 1 or 2. It is a random-effects model, as school effects are defined as random across schools. The Level 1 coefficient was set to zero for all schools. Expressed symbolically, the unconditional model is:

Level 1

$$Y_{ij} = \beta_{0j} + r_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

where Y_{ij} is the outcome for student i in school j; β_{0j} is the intercept, which is school mean achievement; γ_{00} is the grand mean or the average of school means across all schools; r_{ij} is student level residual, representing departure of a student's score from the mean of school j; u_{0j} is school level residual, representing departure of each school's mean from the grand mean.

This unconditional model is in fact a one-way analysis of variance (ANOVA). It partitions variability in the outcome measure into two components: within-school variability and between-school variability. The ratio of between-school variability to the total variability is represented by intraschool correlation, which is the estimated correlation between pairs of scores within schools.

This step is often used as a preliminary check in a hierarchical data analysis to see whether multilevel modeling is appropriate (Bryk & Raudenbush, 1992; Lee, 2000). If intraschool correlation exceeds .1, this indicates that 10% of the total variance in student achievement is due to school differences. Therefore, it is appropriate to use a hierarchical linear model to estimate school effects (Lee, 2000).

Random coefficient model. Following the unconditional model is the random coefficient model that regresses the outcome measure on student characteristics. It provides estimates of variability in the regression coefficients, including the intercept and slopes, across schools. No school level variables are included. In this model, regression coefficients are assumed to vary across the population of schools.

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}\text{PRIORACH} + \beta_{2j}\text{LUNCH} + r_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \mu_{1j}$$

$$\beta_{2j} = \gamma_{20} + \mu_{2j}$$

where Y_{ij} is the outcome measure for student i in school j; β_{0j} is the intercept, which is the adjusted school mean achievement; β_{1j} is the expected change in the outcome measure for a unit change in PRIORACH; β_{2j} is the lunch gap, which is the average difference between the achievement of the students in free or reduced lunch and that of those who are not, in school j, after controlling for the effects of individual student's prior achievement; r_{ij} is the random error, representing a unique effect associated with student *i* in school *j* and $r_{ij} \sim N(0, \sigma^2)$; γ_{00} is the average adjusted school mean across all schools; γ_{10} is the average of PRIORACH-ACHIEVEMENT regression slopes, representing the average effect of prior achievement across all schools; γ_{20} is the average of lunch gap, representing the effects of being in free- or reduced-price lunch program across all schools; u_{ij} is the unique effects of school j on associated parameters.

Intercept- and slopes-as-outcomes model. In the third step, the intercept and slopes from the previous procedure was used as outcome measures and modeled as a function of school characteristics. Please note that the intercept and slopes would be modeled as fixed, random, or nonrandomly varying, depending on the results from the previous step. This model was designed to measure school effects and thus rank schools on residual estimates.

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}\text{PRIORACH} + \beta_{2j}\text{LUNCH} + r_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{LUNCH\%} + \gamma_{02}\text{SUSP\%} + \gamma_{03}\text{TM\%} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\text{LUNCH\%} + \gamma_{12}\text{SUSP\%} + \gamma_{13}\text{TM\%} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}\text{LUNCH\%} + \gamma_{22}\text{SUSP\%} + \gamma_{23}\text{TM\%} + u_{2j}$$

Centering

When the intercept and slopes in level 1 model become outcome variables at level 2, it is important to make sense of these terms. Centering facilitates meaningful interpretations (Bryk & Raudenbush, 1992). Two frequently used centering methods are group mean centering and grand mean centering. Grand mean centering is often used when the researcher is interested in estimating school effects (Philips & Adcock, 1997; Thum & Bryk, 1997). The data were analyzed using computer program HLM 4.04 (Bryk, Raudenbush & Congdon, 1996).

Estimation of School Effects and rank ordering

School effects were estimated using school level residuals. Residuals for the intercept are often used as estimates of school effects. As noted in the above, slopes were also assumed to vary across schools. If the results show slope variability, it indicates school effects interact with related variables. Therefore, residuals for the slopes should be included in the estimation of school effects as well. The estimates of school effects are used to rank order schools.

Results

Results of Principal Components Analysis

Principal components analysis (PCA) of the test scores from the five areas was conducted to derive a composite score as the outcome measure of overall student achievement. Using Kaiser rule of retaining only components whose eigenvalues are greater than 1 (Stevens, 1996), one factor was retained. The one retained factor accounted for 68.92% of the variance in the outcome measure and 70.78% of the variance in prior achievement. Factor loadings were .87 on reading, .85 on mathematics, .89 on science, .92 on citizenship, and .58 on writing for the outcome measure and .86 on reading, .84 on mathematics, .89 on science, .87 on citizenship, and .72 on writing for prior achievement.

Description of the Sample

The final sample included in the study was 1915 sixth-grade students (1999-2000 school year) from 44 schools in the school district. Of these students, 856 (44.70%) were male and 1059 (55.30%) were female; 1201 (62.72%) were in free- or reduced-price lunch program and 714 (37.28%) were not. Students came from one of the six ethnicities: Asian, Black, Hispanic, Indian, Multi-racial, or White. Of them 14 (.73%) were Asians;

822 (42.92%) were Blacks; 134 (7.00%) were Hispanics; 2 (.10%) were Indians; 8 (.42%) were multi-racial; and 935 (48.83%) were Whites.

The number of students included in each school varied from a minimum of 15 to a maximum of 91. Both the outcome measure and prior achievement had a mean of 100 and standard deviation of 15 (minimum = 45 and maximum = 157 for the outcome measure and minimum = 24 and maximum = 155 for prior achievement). Descriptive statistics of the outcome measure and prior achievement by school are provided in tables 1 and 2. Histogram of achievement variables shows that these variables are normally distributed and no extreme values were found. Pearson correlation between the composite outcome measure and prior achievement was .80.

INSERT TABLES 1 AND 2 ABOUT HERE

Descriptive statistics of the schools that hosted the students are displayed in Table 3. The table shows that there is considerable variation among these schools on most of the characteristics except on percentage of minority students (MINORITY%) and percentage of female students (FEMALE%). Correlations are .68 between LUNCH% and SUSP%, –.66 between LUNCH% and TM%, and –.61 between TM% and SUSP%.

INSERT TABLE 3 ABOUT HERE

Results of Hierarchical Linear Modeling

As described above, school effects were estimated following a three-step procedure: unconditional model, random coefficient model, and intercept- and slopes-as-outcomes model. The results of each step using composite scores as the outcome measure are presented successively here.

Unconditional model. Table 4 presents the results from the unconditional model. The maximum likelihood point estimate for the grand mean was 100. 35. The maximum likelihood estimate of within-school variance (δ^2) was 172.67 and that of between-school variance (τ) was 51.82. The intraschool correlation (ρ), represented by the ratio of between-school variance to total variance, was:

$$\rho = \tau / (\tau + \delta^2) = 51.82 / (51.82 + 172.67) = .23$$

Therefore, 23% of the variance in student achievement lies in between-school differences, and, consequently, 77% of it lies in within-school variability.

Table 4

Results from the Unconditional Model

Fixed effect	Coefficient	SE	T-ratio	p-value
Average school mean, γ_{00}	100.35	1.13	88.93	.000
Random effects	Variance component	df	Chi-square	p-value
School mean, u_{0j}	51.82	43	624.19	.000
Level 1 effect, r_{ij}	172.67			

Hierarchical linear modeling also provides reliability estimates. Reliability in this process is the average reliability of using schools' mean to estimate population mean. As sample size varies from school to school, the reliability of using sample means as an indicator of population mean also differs. The reliability estimate, which is .91, shows that sample means were reliable estimates of the population mean.

The result of the Chi-square test ($\chi^2 = 624.19$, $p = .000$) indicates that homogeneity of variance was not tenable across schools. In addition, the intraschool correlation exceeds .10. Therefore, the use of HLM in this study is appropriate.

Random coefficient model. Table 5 presents the results from the random coefficient model. It shows that the average of school means was 99.68. The PRIORACH-achievement slope ($\gamma_{10} = .78$, $p = .000$) indicates that prior composite achievement (fourth grade) correlated positively with composite achievement (sixth-grade) within a school. That is, students' sixth-grade composite score increased, on average, with .78 for one point increase on their fourth-grade composite score. The average lunch gap was -1.33, indicating that students in lunch program scored on average 1.33 points lower than those with similar prior achievement, but not in lunch program. The associated t-ratios and their p-values show that both of the predictors were statistically significant.

Random effects provide estimates of variance and test the hypothesis that the variance was zero across schools. The estimated residual variance among the means was 10.60, with $\chi^2 = 239.93$, $p = .000$, suggesting that there were significant differences among the 44 school means. The estimated variance of the slopes was 1.0E-2, with $\chi^2 = 83.45$, $p = .000$. Therefore the hypothesis that the variance was null was rejected, indicating that the relationship between PRIORACH and achievement within schools varied significantly across the schools. The estimated variance for lunch gap was 2.07, with $\chi^2 = 48.19$, $p = .271$. This indicates that lunch gap did not vary significantly across the 44 schools.

Table 5

Results from the Random Coefficient Model

	Fixed effects			
Outcome variable	Coefficients	SE	T-ratio	p-value
Adjusted mean achievement, γ_{00}	99.68	.54	183.68	.000
Mean PRIORACH-achievement slope, γ_{10}	.78	2.2E-2	35.89	.000
Mean Lunch gap, γ_{20}	-1.33	.53	-2.52	.016
	Random effects			
Outcome variable	Variance	df	χ^2	p-value
Adjusted Mean achievement, u_{0j}	10.60	43	239.93	.000
PRIORACH-achievement slope, u_{1j}	1.0E-2	43	83.45	.000
Lunch gap, u_{2j}	2.07	43	48.19	.271
Level 1 effect, r_{ij}	69.40			

Random coefficient modeling is an important process, as it provides guidance on the final specification of level 1 coefficients using three options: random, fixed, or nonrandomly varying. The reliability estimate is another useful statistic, as it indicates the amount of variation in the coefficients that is potentially explainable by level 2 variables (Bryk & Raudenbush, 1992). The reliability estimate of the intercept is based on the sample size and that of a slope on both the sample size and the variability of the slopes within the school. Reliability for the intercept was relatively higher (.74) than that for prior achievement (.45). However, reliability for LUNCH slope was very low (.16). The lack of precision, or low reliability, of lunch gap was due to the relative homogeneity of its distribution among many of the schools.

The variance estimate at the student level in this random coefficient model was reduced to 69.40, compared to that of 172.67 from the unconditional model reported in Table 4. This proportion of reduced variance was calculated using (172.67- 69.40) / 172.67 = .5981. Thus, student prior achievement and lunch status accounted for 59.81% of the variance in achievement at the student level.

The inclusion of student level variables typically reduces between-school variability. In this step, the maximum likelihood estimate of intra-school variance (δ^2) became 69.40 and that of inter-school variance (τ) became 10.60. The intraschool correlation (ρ) was:

$$\rho = \tau / (\tau + \delta^2) = 10.60 / (10.60 + 69.40) = .13$$

Therefore, with the introduction of PRIORACH and LUNCH, between-school variance is reduced to 13%. Consequently, within-school variance becomes 87%.

After fitting the level 1 model, the assessment of the model assumption is necessary (Bryk & Raudenbush, 1992). Hierarchical modeling assumes that errors at the student level are independent, normally distributed with a mean of zero and constant variance. The homogeneity of variance was tested using the Chi-square statistic, which was statistically significant ($\chi^2 = 77.43$, $\underline{p} = .001$). However, this finding is inconsistent with that from the Chi-square test in the empty model ($\chi^2 = 56.93$, $\underline{p} = .076$). One explanation for the heterogeneity of variance is the non-normal distribution of LUNCH with a heavy tail (62.7% in free or reduced lunch program versus 37.3% not), although common reasons include the omission of an additional important level 1 variable and fixing the effects of a level 1 predictor that is random or nonrandomly varying (Scientific Software International, personal communication, February 23, 2000). Therefore, this violation of homogeneity of variance is not of grave concern here.

The results from the random coefficient model show that student level predictors, fourth-grade achievement and lunch status, on average, had significant associations with sixth-grade achievement. Further, the Chi-square test and reliability indices indicate that there was considerable variability among schools in the adjusted mean achievement and PRIORACH-achievement slope, but not in lunch gap. These results were used for the specification of the final model.

A common practice in school effects studies is to model intercept as random only. However, with respect to the results found thus far, it would be inappropriate to specify the PRIORACH-achievement slope as fixed when significant slop variability was found to exist among the schools. If slope variability is modeled as fixed, the estimates of level 2 coefficients will be biased (Bryk & Raudenbush, 1992). Therefore, both the intercept and PRIORACH-achievement slope should be specified as random and lunch gap as fixed in the final step.

Intercept- and slopes-as-outcomes model. Based on the above procedure, the model at this step becomes:

Level 1

$$Y_{ij} = \beta_{0j} + \beta_{1j}\text{PRIORACH} + \beta_{2j}\text{LUNCH} + r_{ij}$$

Level 2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\text{LUNCH\%} + \gamma_{02}\text{SUSP\%} + \gamma_{03}\text{TM\%} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}\text{LUNCH\%} + \gamma_{12}\text{SUSP\%} + \gamma_{13}\text{TM\%} + u_{1j}$$

$\beta_{2j} = \gamma_{20}$

It is assumed that the intercept and the slope vary not only as a function of the three predictors, LUNCH%, SUSP%, and TM%, but also a function of the unique school effect. Substituting the student level model $Y_{ij} = \beta_{0j} + \beta_{1j}$PRIORACH + β_{2j}LUNCH + r_{ij} with school level equations yields the combined model:

$Y_{ij} = (\gamma_{00} + \gamma_{01}\text{LUNCH\%} + \gamma_{02}\text{SUSP\%} + \gamma_{03}\text{TM\%} + u_{0j}) + (\gamma_{10} +$

$\gamma_{11}\text{LUNCH\%} + \gamma_{12}\text{SUSP\%} + \gamma_{13}\text{TM\%} + u_{1j})\ \text{PRIORACH} + \gamma_{20}\text{LUNCH} + r_{ij}$

Results from the final intercept- and slopes-as outcome model are presented in Table 6. It shows that the average percentage of students in free- or reduced-price lunch in a school was negatively related to adjusted school mean achievement (γ_{01} = -5.5E-2, p = .117), but not statistically significant. The effects of SUSP% were almost negligible (γ_{02} = -1.1E-2, p = .927). The average percentage of teachers with a Master's degree was significantly correlated with adjusted school mean achievement in a negative way (γ_{02} = -1.0E-2, p = .048). This is unexpected. There were two possible explanations concerning this result. First, adjusted school mean achievement was significantly lower in schools with higher percentage of teachers with a Master's degree. Second, adjusted school mean achievement was higher in schools with higher percentage of teachers with a Master's degree and lower in schools with lower percentage of teachers with a Master's degree. Concerning prior achievement, its differential effect on the adjusted mean achievement depends also on the percentage of teachers with a Master's degree. The results show that the percentage of teachers with a Master's degree was positively correlated with prior achievement (γ_{12} = 4.5E-3) and this correlation was statistically significant (p = .047). There are no demonstrated significant effects of SUSP% and LUNCH%.

Table 6

Results from the Intercept- and Slopes-as-Outcomes Model on Composite 2000

	Fixed effects			
Outcome variable	Coefficients	SE	T-ratio	p-value
Model for adjusted school means				
Intercept, γ_{00}	107.29	3.40	31.51	.000
Mean LUNCH%, γ_{01}	-5.5E-2	3.4E-2	-1.60	.117
Mean SUSP%, γ_{02}	-1.1E-2	.12	-.09	.927
Mean TM%, γ_{03}	-1.0E-2	5.0E-2	-2.03	.048
Model for prior achievement				
Intercept, γ_{10}	.61	.15	4.19	.000
Mean LUNCH%, γ_{11}	5.2E-4	1.4E-3	.38	.707
Mean SUSP%, γ_{12}	-2.6E-3	5.1E-3	-.52	.606
Mean TM%, γ_{13}	4.5E-3	2.2E-3	2.05	.047
Model for lunch gap				
Intercept, γ_{20}	-1.33	.48	-2.74	.010

	Random effects			
Outcome variable	Variance	df	χ^2	p-value
Adjusted school mean residual, u_{0j}	9.08	40	220.50	.000
PRIORACH-achievement slope residual, u_{1j}	9.4E-3	40	75.99	.001
Level 1 effect, r_{ij}	69.82			

Percentage of teachers with a Master's degree in this analysis seemingly had mixed effects: a negative one on the adjusted school mean achievement and a positive one on the prior achievement. The amount of level 1 variance explained in this model was 69.82, close to 69.40 reported in the random coefficient model in Table 15. This result is expected, since the same level 1 predictors were included in both models.

Variance explained at level 2 can be estimated using a procedure similar to that of level 1. The school residual variance, $u_{0j} = 9.08$, is the variance not accounted for by the school level variables included in the model. Recall that the unconditional variance of the intercept from the random coefficient model was 10.60, as reported in Table 5. With the school characteristics taken into account, the school mean residual variance became 9.08. Therefore, the variance was reduced by (10.60 - 9.08) / 10.60 = .14. That is, LUNCH%, SUSP%, and TM% accounted for 14% of the variance in adjusted school mean achievement at the school level. Consequently, 86% of the variance in the adjusted school mean achievement was school effects. The test of homogeneity of variance (χ^2 = 220.50, $\underline{p}$ = .000) rejects the null hypothesis that there was no significant residual variance to be explained in the intercept. This indicates that other school-level factors, such as school practice, may account for the remaining part of the variance. With regard to this study, it is in fact part of the school effects.

The residual variance of the PRIORACH-achievement slope is 9.4E-3, showing no substantial reduction from the unconditional variance of 1.0E-2 from Table 5. This suggests that variation among schools in average prior achievement remains unexplained. The Chi-square test was statistically significant (χ^2 = 75.99, $\underline{p}$ = .001), thus rejecting the null hypothesis that the residual variance of the slope was zero. As both the intercept and the slope were specified as random effects, their residuals are interpreted as effects associated with each school. Thus, the significance of the residual variance of both parameters indicates heterogeneity among the estimates of schools effects.

The computer software HLM 4.04 provides empirical Bayes (EB) residuals, ordinary least squares (OLS) residuals, and fitted values for the intercept and slopes. Level 1 coefficients are often estimated using the empirical Bayes procedure (Philips & Adcock, 1997). The empirical Bayes estimation produces "optimal composites" of an estimate on the basis of the data from a school and also by borrowing information from other similar schools (Bryk, Raudenbush, & Congdon, 1996). Empirical Bayes residuals are often used as measures of individual school performance, as "they take into account group membership when the number of groups is large and produce relatively stable estimates even when sample sizes per school are modest" (Bryk & Raudenbush, 1992, p. 124). They are also called "shrunken" residuals as its estimated variance tends to be less than the estimated true variance and that the residuals are "shrunk" towards the mean.

Estimation of School Effects

A school effects study with slope variability is characterized as differential effectiveness, as school effects differ on the different values on the slope instead of being uniform across all the students in a school. For instance, this study found the existence of significant variability of PRIORACH slope. This indicates school effects vary across students with different prior achievement. One school may be more effective with students with high prior achievement scores while another may be more effective with students with average prior achievement. Although differential effectiveness captures and characterizes this study, the focus is on its application. Further elaboration can be found in other research (e.g., Aitkin & Zuzovsky, 1994).

As there is significant variability among the schools on prior achievement slope, it is inadequate and biased to use a single school level residual term as an indicator of school effects. Recall that the combined model is:

$$Y_{ij} = (\gamma_{00} + \gamma_{01}\text{LUNCH\%} + \gamma_{02}\text{SUSP\%} + \gamma_{03}\text{TM\%} + u_{0j}) + (\gamma_{10} + \gamma_{11}\text{LUNCH\%} + \gamma_{12}\text{SUSP\%} + \gamma_{13}\text{TM\%} + u_{1j})\ \text{PRIORACH} + \gamma_{20}\text{LUNCH} + r_{ij}$$

Following the equation proposed by Pituch (1997), school effects are estimated using $u_{0j} + u_{1j}$* PRIORACH, which indicates the effects of a school on student achievement after controlling for student characteristics and school contextual factors. The residuals, u_{0j} and u_{1j}, are multivariate normally distributed with means of zero, with u_{0j} representing the residual for the intercept and u_{1j} the residual for the PRIORACH-achievement slope. Their EB residual estimates of the intercept and the PRIORACH slope for each school are presented in Table 7. The above equation also indicates that school effects depend on student's prior achievement. When the slope residual (u_{1j}) for a particular school is zero, its effects are constant and the same estimates of effects are obtained for the school, regardless of prior achievement. When the slope residual is not zero, it indicates the existence of the interaction between school effects and student prior achievement. School effects vary on different prior achievement scores. Different effects may result from different values on prior achievement.

INSERT TABLE 7 ABOUT HERE

Three values on prior achievement were selected for the estimation of school effects. They were the mean (PRIORACH = 100), one standard deviation below the mean (PRIORACH = 85) and one standard deviation above the mean (PRIORACH = 115). A school may have three different estimates of effects and thus possibly three rankings depending on the estimates.

Estimates of school effects on three values of prior achievement and their rankings are presented in Table 8. Take School 21290 for example. The residual estimate of u_{0j} is 3.8731 and that of u_{1j} is .0755, as shown in Table 7. Thus, its school effects are 3.8731 + .0755*PRIORACH. Substituting PRIORACH with three values yields three estimates of school effects: 10.29, 11.42 and 12.56, as reported in Table 8. This means that for 21290 its school effects are 10.29 for students scoring one standard deviation below the mean on prior achievement, 11.42 for those with average prior achievement and 12.56 for those scoring one standard deviation above the mean on prior achievement. These estimates also show that the effects of this school increase with the increase of student prior achievement, as u_{1j} is positive. Even though the estimates of effects are different, the school's ranking does not change across three prior achievement values (ranking 6). Take a look at another school, 11040. Its u_{0j} and u_{1j} are -.4491 and -.0492, respectively, as shown in Table 7. Its school effects are -.4491 + (-.0492*PRIORACH), thus yielding an estimate of –4.63 for students with prior achievement equal to one standard deviation below the mean, –5.37 for students with average prior achievement and –6.11 for those with prior achievement being one standard deviation above the mean. The effects of this school decreased with the increase of students' prior achievement, as its u_{1j} is negative.

INSERT TABLE 8 ABOUT HERE

In general, the results show that most of the school rankings remain relatively stable across three prior achievement scores, 30 of the schools (68%) retaining the same ranking, and 14 (32%) of them fluctuating within three ranks. This indicates that these schools do not vary dramatically in terms of their effects on students with prior achievement falling in the middle 68%, i.e., one standard deviation around the mean. There may be greater discrepancy among school effects on students with extreme prior achievement. They tend to be comparatively constant for students around average prior achievement. This also suggests that the impact of variability of the residuals for the intercept predominates that for the slope.

Discussion

The findings of the study show that lunch status, as a measure of student SES, is significantly related to student achievement. This is consistent with findings documented in literature. Low SES has a strong negative correlation with all educational outcomes (Saunders, 1997). The results of this study provide further evidence in support of this association. However, the percentage of students in free- or reduced-price lunch, as a measure of school SES, is not associated with achievement, after controlling for the percentage of students suspended and the percentage of teachers with a Master's degree. This is inconsistent with the findings of Caldas and Bankston III (1997) and Goldstein (1997), who noted that attending schools with classmates from families of higher SES increases one's own academic performance independent of one's own SES status, race, and other factors. Similarly, the percentage of students suspended does not associate with achievement.

The effects of the percentage of teachers with a Master's degree on achievement are mixed. It is found that the percentage of teachers with a Master's degree is associated negatively with sixth-grade achievement, but positively with fourth-grade achievement. One possible reason might be the limited sample size of schools.

The study was designed to measure value added school effects on Ohio Sixth-Grade Proficiency Test results. The value added conceptualization defines effectiveness as a measure of school effects after controlling for intake characteristics, including prior attainment and student characteristics. One of the major findings of this study is that prior achievement is a significant predictor of current achievement. That is, students' fourth-grade OPT scores predict significantly their sixth-grade performance. In addition, this relationship between prior achievement and achievement varies significantly across all the schools. This indicates that a school effects study without using proper prior achievement is not feasible, thus providing further evidence in support of using value added measure to assess school or district performance.

The results of the study also show that the use of hierarchical linear modeling for measuring school effects is appropriate. When no variables are included in the model, intraschool correlation exceeds .10, indicating that between-school differences are not trivial. They account for 23% of the total variance in the outcome measure. Systematic differences exist among schools and school effects are not negligible.

The proportion of variance that can be explained by school factors when student characteristics are taken into account is also identified. Between-school variability accounts for 13% of the variance in composite student achievement, after controlling for prior achievement and lunch status. These findings fall within the range reported in literature. Creemers (1994) summarized the results of school effectiveness research and concluded that school and classroom factors explained 12 - 18% of the variance in student achievement after taking student background into account. Bosker and Witziers (1996) reported that school level factors accounted for 10% of the variance in student achievement after adjusted for student background in a sample of U.S. studies using hierarchical linear modeling.

One limitation of the study is the concept value added is inadequate. Value added implies that student progress is entirely attributed to school, although only part of it should be in reality (e.g., in the absence of intervention programs) (Fitz-Gibbon & Kochan, 2000). In addition, value added measure is technically represented by school level residuals. In this case, it is difficult to know how much of the between-school variance is accounted for by random error and how much by stable and systematic properties of school (Rowan, Bossert, & Dwyer, 1983).

This study shows that student prior attainment predicts significantly later achievement. Scores unadjusted for student intake characteristics do not provide a fair and equitable measure of school effects. Therefore, it is highly recommended that a value added rating system using prior achievement be used.

One of the findings of the study is that school effects change with different student prior achievement. Three values around the mean of prior achievement were selected to compute school effects. It may be worthwhile to investigate the change of school effects using other different prior achievement scores.

This study investigates school effects using two-level hierarchical linear modeling. Variables at each level are included independently at each level, without modeling the interactions between variables. Future studies may incorporate these interactions into the model.

Effective schools can also be identified from the results of this study. The purpose of identifying effective schools is to identify effective school characteristics so as to disseminate them within similar schools and enhance the quality of public education. Therefore, future study should be conducted to identify effective characteristics, such as school practices that are beneficial to promoting academic achievement. A qualitative study or a combination of quantitative and qualitative study can serve best such a purpose.

References

Aitkin, M., & Zuzovsky, R. (1994). Multilevel interaction models and their use in the analysis of large-scale school effectiveness studies. School effectiveness and school improvement, 5, 45-73.

Austin, G., & Reynolds, D. (1990). Managing for improved school effectiveness: An international Survey. School Organization, 10(2-3), 167-178.

Bosker, R, J., & Witziers, B. (1996). The magnitude of school effects, or: Does it really matter which school a student attends? Paper presented at the annual meeting of the American Educational Research Association, New York: NY.

Brown, S., Duffield, J., & Riddell, S. (1997). School effectiveness research: the policy makers' tool for school improvement? In A. Harris, N. Bennett and M. Preedy (Eds.) Organizational effectiveness and improvement in education. (pp. 138-146). Buckingham, Philadelphia: Open University Press.

Bryk, A., Raudenbush, S., & Congdon, R. (1996). Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs. Chicago, IL: Scientific Software International, Inc.

Bryk, A., & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: SAGE.

Caldas, S. J. (1993). Reexamination of input and process factor effects on public school achievement. Journal of Education Research, 84(4), 206-214.

Caldäs, S. J., & Bankston III, C. (1997). Effect of school population socioeconomic status on individual student academic achievement. The Journal of Educational Research, 90(5), 269-277.

Coe, R., & Fitz-Gibbon, C. T. (1998). School effectiveness research: Criticisms and recommendations. Oxford Review of Education, 24(4), pp. 421-438.

Cohen, M. (1983). Instructional management and social conditions in effective schools. In A. O. Webb & L.D. Webb (Eds.). School finance and school improvement: Linkage in the 1980's (pp. 17-51). Cambridge, MA: Balinger.

Creemers, B. P. M. (1994). The history, value and purpose of school effectiveness studies. In D. Reynolds, B. P. M. Creemers, P. S. Nesselrodt, E. C. Schaffer, S. Stringfield, & C. Teddlie (Eds.), Advances in school effectiveness research and practice (pp.9-23), Tarrytone, NY: Elsevier Science.

Crone, L. J., Lang, M. H., Franklin, B. J., & Halbrook, A. M. (1994). Composite versus component scores: consistency of school effectiveness classification. Applied Measurement in Education, 7(4), 303-321.

Educational Testing Service (ETS). (1999). Order in the classroom: Violence, discipline, and student achievement. ETS policy information report. Princeton, NJ: Educational Testing Service.

Fitz-Gibbon, C., & Kochan, S. (2000). School effectiveness and education indicators. In T. Charles & D. Reynolds (Eds.), The International Handbook of School Effectiveness Research (pp. 257-282). London, New York: Falmer Press.

Frederick. J. M. (1987, January). Measuring school effectiveness: Guidelines for educational practitioners (ERIC/TME Report 93). ERIC Clearinghouse on Tests, Measurements, and Evaluation. Educational Testing Service, Princeton, NJ.

Goldstein, H. (1995). Multilevel statistical models. New York: Halsted Press.

Goldstein, H. (1997). Methods in school effectiveness research. School Effectiveness and School Improvement, 8(4), 369-395.

Good, T. L., & Weinstein, R.S. (1986). Schools make a difference. American Psychologist, 41(10), 1090-1097.

Hanushek, E. A. (1992, Februrary). The trade-off between child quantity and quality. The Journal of Political Economy, 100(1), 84-117.

Haycock, K. (1998, Fall). Good teaching matters ... a lot. OAH Magazine of History, 13(1), 61-63.

Holdaway, E. A., & Johnson, N. A. (1993, August). School effectiveness and effectiveness indicators. School Effectiveness and School Improvement, 4(3), 165-188.

Johnson, R.A., and Wichern, D.W. (1998). Applied Multivariate Statistical Analysis (4th ed.). Upper Saddle River, NJ: Prentice Hall, Inc.

Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. Educational Psychologists, 35(2), 125-141.

Levine, D. U., & Lezotte, L. W. (1990). Unusually effective schools: A review and analysis of research and practice. Madison, WI: National Center for Effective Schools.

Mandeville, G. K., & Heidari, K. (1988). Measuring school effectiveness using hierarchical linear models. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

McPherson, A. (1997). Measuring added value in schools. In A. Harris, N. Bennett and M. Preedy (Eds.) Organizational effectiveness and improvement in education (pp. 184-190). Buckingham, Philadelphia: Open University Press.

Mendro, R. L., Jordan, H. R., Gomez, E., Anderson, M. C., & Bembry, K. L. (1998). An application of multiple linear regression in determining longitudinal teacher effectiveness. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Odden, A., & Webb, L. D. (1983). The linkages between school finance and school improvement. In A. O. Webb & L.D. Webb (Eds.), School finance and school improvement: Linkage in the 1980's (pp. xiii-xxi). Cambridge, MA: Balinger.

Ohio Department of Education. (2000). Accountability and improvement Senate Bill 55 [CD- ROM]. State of Ohio 2000 District, Building and State Report Cards. Columbus, OH: Ohio State Department of Education.

Pituch, K. A. (1999, April). Describing school effects with residual terms. Evaluation Review, 23(2), pp. 190-211.

Phillips, G. W., & Adcock, E. P. (1997). Measuring school effects with hierarchical linear modeling: Data handling and modeling issues. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. Elementary School Journal, 83(4), 427-52.

Rachor, R. E. (2000). A comparison of the most advantaged with the least advantaged school districts. Unpublished study.

Reynolds, D., & Teddlie, C. (2000). The future agenda for school effective research. In T. Charles & D. Reynolds (Eds.), The International Handbook of School Effectiveness Research (pp. 322-343). London, New York: Falmer Press.

Rowan, B., Bossert, S., & Dwyer, D. C. (1983, April). Research on effective schools: A cautionary note. Educational Researcher, 12(4), 24-31.

Sammons, P., Mortimore, P., & Thomas, S. (1996). Do schools perform consistently across outcomes and areas? In J. Gray, D. Reynolds, C. Fitz-Gibbon and D. Jesson (Eds), Merging traditions: The future of research on school effectiveness and school improvement (pp. 3-29). London: Cassell.

Saunders, L. (1997). Value-added principles, practice and ethical considerations. In A. Harris, N. Bennett and M. Preedy (Eds.) Organizational effectiveness and improvement in education. (pp. 191-202). Buckingham, Philadelphia: Open University Press.

Sanders, W. L. (1998, December). Value-added assessment. School Administrator, 55(11), 24-27.

Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student academic achievement. Research Progress Report.

Sanders, W. L., Wright, S. P., Horn, P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. Journal of Personnel Evaluation in Education, 11(1), 57-67.

Schagen, I. (1991, Winter). Beyond league tables: How modern statistical methods can give a truer picture of the effects of schools. Educational Research, 33(3), 216-22.

Scheerens, J., & Bosker, R. (1997). The foundations of educational effectiveness. Oxford, OX; New York, NY: Pergamon.

Stevens, J. (1996). Applied multivariate statistics for the social science (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Stoll, L., & Fink, D. (1996). Changing our schools: Linking school effectiveness and school improvement. Buckingham, Philadelphia: Open University Press.

Strand, S. (1997). Pupil progress during Key Stage 1: A value added analysis of school effects. British Educational Research Journal, 23(4), 471-487.

Stone, J. E. (1999). Value-added assessment: An accountability revolution. In M. K. & C. E. Finn, Jr (Eds.), Better teachers, better schools (pp. 239-249). Washington, D. C.: Thomas B. Fordham Foundation; Education Leaders Council.

Teddlie, C., Reynolds, D., & Sammons, P. (2000). The methodology and scientific properties of school effectiveness research. In T. Charles & D. Reynolds (Eds.), The International Handbook of School Effectiveness Research (pp. 55-133). London, New York: Falmer Press.

Thomas B. Fordham Foundation (1999). The teachers we need and how to get more of them: A manifesto. In M. K. & C. E. Finn, Jr (Eds.), Better teachers, better schools (pp. 1-18). Washington, D. C.: Thomas B. Fordham Foundation; Education Leaders Council.

Thum, Y. M., & Bryk, A. S. (1997). Value-added productivity indicators: The Dallas system. In J. Millman (ed.) Grading teachers, grading schools: Is student achievement a valid evaluation measure? (pp.100-109). Buckingham, Philadelphia: Open University Press.

Thrupp, M. (1999). School making a difference: Let's be realistic. Buckingham, Philadelphia: Open University Press.

Webster, W. J., & Mendro, R. T., & Almaguer (1993). Effective indices: The major component of an equitable accountability system. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA.

Willms, J. D. (1992). Monitoring school performance: A guide for educators. Washington, D.C. London: The Falmer Press.

Witte, J. F. & Walsh, D. J. (1990, Summer). A systematic test of the effective schools model. Educational Evaluation and Policy Analysis, 12(2), 188-212.

Table 1

Descriptive Statistics of Outcome Measure by School

School	N	M	Median	SD
11000	59	105.38	101.92	15.45
11020	39	114.75	116.81	11.43
11030	27	96.16	94.82	10.75
11040	46	106.79	107.93	10.90
11050	23	99.00	100.97	13.23
11060	39	95.97	93.44	11.22
11070	31	111.34	111.89	12.35
11090	47	103.61	103.54	11.87
11100	15	102.29	101.31	11.35
11110	39	114.73	114.72	12.94
11120	18	105.98	105.05	11.34
11140	30	113.29	112.57	13.45
11150	43	93.07	94.13	12.69
11160	30	95.77	93.43	13.05
11190	50	93.34	93.85	15.22
21210	61	93.94	91.74	11.37
21230	30	110.95	112.43	15.90
21240	54	108.78	110.90	13.57
21270	64	100.75	101.66	11.86
21290	26	98.16	93.50	14.03
31300	32	98.05	98.45	13.10
31310	25	95.07	95.58	9.55
31320	97	110.93	111.48	13.67
31340	66	97.60	97.72	13.41
31350	39	101.26	101.82	11.80
31360	38	92.86	93.74	15.58
31380	42	99.29	97.66	13.11
31390	33	98.98	98.42	11.42
41400	46	106.39	108.66	13.73
41410	52	103.57	100.48	18.17
41420	16	110.19	110.67	10.39
41450	61	89.89	89.18	10.29
41480	54	101.53	100.76	14.82
41490	46	102.29	101.60	13.42
51500	57	98.35	96.57	12.26
51520	48	88.23	87.56	13.90
51540	59	89.85	89.08	11.69
51560	68	88.65	88.73	13.86
51570	49	86.58	85.11	12.57
51590	53	103.10	100.57	14.31
61600	16	95.24	96.64	14.58
61620	33	91.58	92.04	13.05
61630	91	98.56	97.83	12.67
61820	23	105.63	104.93	12.92

Table 2

Descriptive Statistics of Prior Achievement by School

School	N	M	Median	SD
11000	59	107.31	106.50	13.23
11020	39	121.29	119.70	14.65
11030	27	100.20	99.93	9.07
11040	46	110.48	110.67	10.78
11050	23	95.82	96.34	14.06
11060	39	94.15	93.13	11.57
11070	31	111.87	114.07	10.54
11090	47	100.87	99.79	11.03
11100	15	110.94	110.43	10.22
11110	39	108.22	106.28	11.76
11120	18	109.27	107.91	10.01
11140	30	106.47	107.11	10.38
11150	43	89.00	90.77	12.43
11160	30	101.77	99.83	11.50
11190	50	89.53	90.69	12.02
21210	61	92.80	92.52	12.32
21230	30	112.32	114.07	12.33
21240	54	112.82	113.47	12.15
21270	64	96.06	94.62	11.64
21290	26	92.36	90.59	13.46
31300	32	98.23	97.22	13.81
31310	25	96.90	96.17	8.33
31320	97	107.69	107.77	13.28
31340	66	101.69	101.67	10.38
31350	39	98.76	96.90	12.87
31360	38	99.53	98.75	14.70
31380	42	100.84	100.50	18.18
31390	33	98.10	99.24	10.52
41400	46	101.02	100.02	13.00
41410	52	104.28	103.27	16.51
41420	16	110.29	112.44	9.85
41450	61	88.44	87.19	10.18
41480	54	101.38	101.97	15.33
41490	46	99.62	97.58	12.94
51500	57	102.17	99.99	10.30
51520	48	89.41	92.55	16.48
51540	59	91.19	91.99	12.78
51560	68	87.72	87.56	15.43
51570	49	85.25	87.54	17.15
51590	53	102.53	101.23	13.65
61600	16	92.43	91.60	13.51
61620	33	96.67	96.59	11.03
61630	91	101.83	101.08	12.13
61820	23	103.10	102.79	12.70

Table 3

Descriptive Statistics of School

School Characteristics*	Minimum	Maximum	M	SD
SIZE	215.00	889.00	499.41	163.50
LUNCH%	13.10	100.00	69.56	25.68
MOBILITY%	9.40	60.10	31.41	11.57
SUSP%	0.00	24.30	10.70	6.90
TM%*	17.00	68.00	36.50	14.01
AVERT	9.00	30.00	16.82	4.62
BLACK%	1.45	97.18	43.70	34.15
HISPANIC%	0.18	23.85	6.69	6.18
MINORITY%	0.00	3.99	1.61	1.03
WHITE%	1.90	96.40	48.00	31.41
FEMALE%	42.09	54.17	48.33	2.81

Note. All the school characteristics except school size and average years of teaching are expressed in percentage, which is the ratio of the students in such a category to the total number of students in the school. TM% refers to the percentage of teachers with a Master's degree. AVERT is the average years of teaching.

Table 7

EB Residual Estimates of the Intercept and the PRIORACH Slope

School	Intercept (u_{0j})	Slope (u_{1j})
11000	-1.6965	0.0397
11020	-2.5541	-0.1195
11030	-2.5336	-0.0626
11040	-0.4491	-0.0492
11050	2.5635	0.0356
11060	0.3635	-0.0111
11070	-0.0082	0.0313
11090	3.1620	0.0920
11100	-5.0638	-0.1124
11110	5.0405	0.0739
11120	-0.8477	-0.0212
11140	6.5721	0.1712
11150	1.3034	-0.0168
11160	-2.4332	-0.0001
11190	2.2430	0.1413
11210	-1.8731	-0.0946
21230	-0.3611	0.0312
21240	-0.8158	-0.0100
21270	3.0128	0.0485
21290	3.8731	0.0755
31300	0.3753	-0.0688
31310	-1.1605	-0.0298
31320	1.9566	0.0447
31340	-0.1706	0.0534
31350	1.5640	0.0032
31360	-3.2779	0.0194
31380	-1.2307	-0.1090
31390	1.4967	0.0536
41400	4.7022	0.0804
41410	-0.2523	0.0520
41420	-0.3506	0.0026
41450	-1.1028	-0.0736
41480	-0.1839	0.0177
41490	2.9766	0.0649
51500	-2.9171	-0.0323
51520	-3.8339	-0.0913
51540	-3.1867	-0.0483
51560	-2.2877	-0.0800
51570	-2.5342	-0.1620
51590	2.1951	0.0285
61600	2.0314	0.0464
61620	-2.8801	-0.0210
61630	-3.0491	-0.0263
61820	1.6225	0.0333

Table 8

Estimates of School Effects Using PRIORACH = 85, 100, and 115 for Composite 2000

School	Effects1	Rank1	Effects2	Rank2	Effects3	Rank3
11000	1.68	20	2.27	19	2.87	19
11020	-12.71	42	-14.50	42	-16.30	42
11030	-7.85	37	-8.79	37	-9.73	37
11040	-4.63	30	-5.37	31	-6.11	32
11050	5.59	12	6.12	12	6.66	12
11060	-.58	24	-.75	24	-.91	24
11070	2.65	17	3.12	17	3.59	17
11090	10.98	5	12.36	5	13.74	4
11100	-14.62	43	-16.30	43	-17.99	43
11110	11.32	4	12.43	4	13.54	5
11120	-2.65	28	-2.97	28	-3.29	28
11140	21.12	1	23.69	1	26.26	1
11150	-.12	22	-.38	23	-.63	23
11160	-2.44	27	-2.44	27	-2.44	27
11190	14.25	2	16.37	2	18.49	2
11210	-9.91	39	-11.33	39	-12.75	39
21230	2.29	18	2.76	18	3.23	18
21240	-1.67	26	-1.82	26	-1.97	26
21270	7.14	8	7.86	8	8.59	8
21290	10.29	6	11.42	6	12.56	6
31300	-5.47	33	-6.50	34	-7.54	34
31310	-3.69	29	-4.14	29	-4.59	29
31320	5.76	11	6.43	11	7.10	11
31340	4.37	15	5.17	13	5.97	13
31350	1.84	19	1.88	20	1.93	20
31360	-1.63	25	-1.34	25	-1.05	25
31380	-10.50	40	-12.13	40	-13.77	40
31390	6.05	9	6.86	9	7.66	9
41400	11.54	3	12.74	3	13.95	3
41410	4.17	16	4.95	16	5.73	14
41420	-.13	23	-.09	22	-.05	22
41450	-7.36	36	-8.46	36	-9.57	36
41480	1.32	21	1.59	21	1.85	21
41490	8.49	7	9.47	7	10.44	7
51500	-5.66	34	-6.15	33	-6.63	33
51520	-11.59	41	-12.96	41	-14.33	41
51540	-7.29	35	-8.02	35	-8.74	35
51560	-9.09	38	-10.29	38	-11.49	38
51570	-16.30	44	-18.73	44	-21.16	44
51590	4.62	13	5.05	14	5.47	15
61600	5.98	10	6.67	10	7.37	10
61620	-4.67	31	-4.98	30	-5.30	30
61630	-5.28	32	-5.68	32	-6.07	31
61820	4.45	14	4.95	15	5.45	16

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC
Educational Resources Information Center

REPRODUCTION RELEASE

(Specific Document)

TM035354

I. DOCUMENT IDENTIFICATION:

Title: Measuring Value Added School Effects on Ohio Sixth-Grade Proficiency Test Using Two-Level Hierarchical Linear Modeling

Author(s): Lei Yu and Donald White

Corporate Source: University of Toledo

Publication Date: 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Lei Yu
Printed Name/Position/Title: Lei Yu Assoc. Measurement Statistician
Organization/Address: ETS Rosedale Road, MS 15L Princeton, NJ 08541
Telephone: 609-683-2393
FAX: 609-683-2130
E-Mail Address: lyu@ets.org
Date: 11/11/03

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org